

Aberystwyth University

An evaluation of the genetic relationships between the Hill Dartmoor and the registered Dartmoor Pony Breed

Hegarty, Matthew; McElhinney, Nicola; Ham, Emily Rose; Potter, Charly; Winton, Clare Louise; McMahon, Robert

Publication date:
2017

Citation for published version (APA):

Hegarty, M., McElhinney, N., Ham, E. R., Potter, C., Winton, C. L., & McMahon, R. (2017). *An evaluation of the genetic relationships between the Hill Dartmoor and the registered Dartmoor Pony Breed*.

Document License
CC BY

General rights

Copyright and moral rights for the publications made accessible in the Aberystwyth Research Portal (the Institutional Repository) are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the Aberystwyth Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the Aberystwyth Research Portal

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

tel: +44 1970 62 2400
email: is@aber.ac.uk

Report prepared for the Friends of the Dartmoor Hill Pony – Dec 2017

An evaluation of the genetic relationships between the Hill Dartmoor and the registered Dartmoor Pony Breed

**Matt Hegarty, Nicola McElhinney, Emily Ham, Charly Morgan, Clare Winton and
Rob McMahon**

Authorised Copy for Release



**Dr. Matt Hegarty
22nd December 2017
Revision A: 2nd February 2018**

Revision Notes:

Rev. A: Removed sentence in section 1.4.1. which incorrectly implied Registered Dartmoor stud book opened and closed in same year.

1.1. Executive Summary

The Dartmoor National Park is home to ponies free-living alongside cattle and sheep. In 2015, Natural England requested evidence supporting the claimed value of these ponies to the biodiversity, tradition and culture of Dartmoor. No evidence existed of whether the Hill Dartmoors are sufficiently distinct from the eleven other recognised pony breeds native to the UK and Ireland.

Preliminary data from a small collection of Hill Dartmoors using genetic analysis suggested that they could be distinguished genetically from the registered Dartmoor breed. We subsequently suggested that we be supplied samples from as broad a cohort of the Hill Dartmoors as possible (25 animals per herd across five herds) for a more conclusive assessment. In actuality, due to time constraints at roundups, we were supplied 105 Hill Dartmoor samples and a further 16 registered Dartmoors (to supplement our existing data). These animals were tested for genetic variation across 15 highly variable simple sequence repeat (SSR) sites and 172 single nucleotide polymorphisms (SNPs). The results were compared to a body of data obtained from previous work at IBERS including the Welsh Section A-D, the free-living Carneddau Hill Pony and other mixed groups of pedigree and non-pedigree UK ponies (Winton *et al.*, 2013).

Results of genetic analysis show that, whilst genetically closer to the registered Dartmoor than to other breeds, shows a genetic distinctiveness that enables assignment of most animals to a Hill Dartmoor grouping when employing SNP markers. When using SSR markers, the lower resolution means that ~50% of the animals are misassigned to other breeds, however misassignment of Hill Dartmoors as registered Dartmoors (or *vice versa*) is rare. We note that high levels of misassignment are also seen in other breeds when employing SSR markers and that, in some cases, misassignment is due to a small number of herds known to have historically bred with Shetland to produce ponies ideal in size and temperament for use as pit ponies.

The Hill Dartmoor is, as might be expected for a free-living population, much more genetically diverse than the other breeds we examined (the Carneddau, whilst also free-living, are a small population of <300 animals and show signs of inbreeding). We would argue that the presence of a common genetic signature across the population despite this adds to the case for recognising the Hill Dartmoor as a separate group to the registered Dartmoor. We also detect a number of SNP markers displaying significant differences in allele frequency between the two groups, some of which are hypothetically linked to trait differences (although further work is required to state this definitively).

1.2. Why this genetic analysis was commissioned

The charity Friends of the Dartmoor Hill Pony commissioned our team to undertake genetic analysis of Dartmoor Hill Ponies (Hill Dartmoors). In 2015 Natural England

requested evidence to recognise the true value of these semi-wild ponies from Dartmoor. This has raised awareness of the ponies and the need to ensure their protection as a rare and valuable ancient type of pony with special abilities to survive in the harsh environment of the moor, whilst creating habitats for rare flora and fauna in the way that they live and graze. However, no evidence existed of whether the Dartmoor Hill Pony are sufficiently distinct from the eleven other recognised pony breeds native to the UK and Ireland, although locally it is noted that other ponies do not thrive when put out on the open moor.

In commissioning our study, Friends of the Dartmoor Hill Pony sought answers to the following questions:

i) Are Dartmoor Hill Ponies one distinct genetic type of pony?

The ponies in the semi-wild herds, roaming free in natural herds on Dartmoor, differ to each other in height and colour. They have not been subject to selection of characteristics by a breed society body. Horse breed registries rely on genetic testing as part of the process to ensure pedigree integrity. Testing has been done on horses for many years to show their genetic variability and determine if they could be on the breed register or not. Horse blood typing tests used to be the most common between 1960 and 1985. At the time these were the most accurate methods available but there were some limitations (Bowling, 2001). Now breed registries can take advantage of genetic markers such as SSRs and SNPs. *Would similar genetic testing indicate that Dartmoor Hill Ponies are one single genetic population, or not?*

To what degree are they distinct from, or similar to, other populations of native pony breeds? To what degree are they distinct from their nearest cousin, the registered Dartmoor Pony?

ii) As the number of Dartmoor Hill Ponies decreases, are we losing valuable genetics?

Dartmoor Hill Ponies live as free-roaming, wild herds on the uplands on Dartmoor. As they are owned and have been present on the moor for centuries, DEFRA defines them as 'semi-wild'. Ponies within the semi-wild herds were selected not by man but by the ability to survive and thrive on the harshest parts of Dartmoor. *Is this ability seen in their genetics?*

1.3. Sampling Methodology

Given their joint ancestral history, there was particular interest in knowing to what degree the genetics of the Dartmoor Hill Pony are distinct or similar to the registered Dartmoor Pony. Based on previous inclusion of six Hill Dartmoor ponies in our phylogenetic work for the Welsh Pony and Cob Society (WPCS), we observed that the Hill Dartmoors could potentially be distinguished from the registered Dartmoor Pony breed using genetic markers such as SSRs and SNPs. We contracted with Friends of the Dartmoor Hill Pony to test ~125 samples from the major herds to determine if this was truly the case.

Friends of the Dartmoor Hill Pony collected the requested samples from Hill Dartmoors. The charity also collected samples from registered Dartmoors, supplementing those already held at Aberystwyth University. In every case, permission was secured from the ponies' owner.

Dartmoor Hill Pony Samples

The desired sample set was to collect from 20 semi-wild herds, spanning as broad an area of the moor as possible, taking from 5 ponies from each herd. The achieved sample set was 105 samples, from male and female ponies, associated with the following farms:

Merrypit,
Sherril,
Cator,
Corndonford (mixed sources).
Foxworthy,
Shapley,
Harford,
Runnage,
Bittleford,
Vixen Tor,
Shallowford,
Leightown,
Lizwell,
Welstor,
Dunstone,
Leightor,
Nattadon,
Upperton,
Collaton

Dartmoor Pony Samples

Samples for analysis collected from 15 pedigree Dartmoor Ponies, male and female, registered by 10 Dartmoor Pony breeders, using the following prefixes:

Shilstone Rocks,
Huttons Ambo,
Vean,
Newoak,
Lickfolds,
Lizwell,
Hisley,
Easdon,
Blackertor,
Collstone
Merribridge

Family trees were inspected. Samples from closely-related Dartmoor Ponies were rejected. A diverse selection of bloodlines are represented.

These samples were added to the library of native pony samples held in our laboratory and previously used for the work published in Winton *et al.* (2013) or a recent project with the Welsh Pony and Cob Society (WPCS).

1.4. Background

The free-living herds of ponies within the Dartmoor National Park are valuable to the eco-system of the park as well as the culture and heritage (Dartmoor National Park, 2017)

1.4.1 Brief History

Wild pony herds, farmed by man, are known to have existed in the middle Bronze Age (c1350BC) on Dartmoor through the discovery of hoof prints during an archaeological dig at Shaugh Prior on the south of the moor (Plymouth Museum). Written evidence from the 11th century onwards evidence a continuous existence on Dartmoor since then, and probably earlier (Friends of the Dartmoor Hill Pony, 2016).

In the 1860s two distinct breeding routes diverged: some keepers of pony herds on Dartmoor, under the influence of the Prince of Wales (a major land-owner on Dartmoor), entered into 'improving' selected ponies from their moor herds by 'judicious outcrossing' (Palmer, 1990) with Arabs, Hackneys and Thoroughbred, Welsh, Exmoor and Fell. This was undertaken in pursuit of producing a pony of beauty, ideal as a sport pony for the new, fashionable sport of polo or as an army pack pony. From this foundation stock, in 1925, the stud book for the registered Dartmoor Pony was opened. The newly formed Dartmoor Pony Breed Society set the characteristics of this new breed – colour, height, conformation of beauty – and further judicious outcrossing achieved the selected ponies that were registered in the stud book. No longer running wild on the moor, the Dartmoor Pony became a properly recorded breed, with breed qualities now governed by human constraints, for example the ponies cannot exceed 12.2hh and cannot be piebald or skewbald (Dartmoor Pony Society, 2016). The selective in-line breeding was under expert and dedicated direction to create the registered Dartmoor Pony of today.

Ponies within the semi-wild herds were selected not by humans but by the ability to survive and thrive on the harshest parts of Dartmoor: the ponies' qualities were governed and selected by the moor itself. Dartmoor is a very diverse area with a great difference in altitude and terrain. Pony herds have adapted to their leared areas: the smaller type thrives on the highest parts of the moor, and on the lower, less harsh areas there are larger pony types.

1.4.2. Reduction in Dartmoor Hill Pony numbers

Numbers of semi-wild pony herds on Dartmoor, and the number of ponies within them, have decreased dramatically over the last 30 years. It is said that there used to be approximately 30,000 ponies running on the commons of Dartmoor, now there is estimated to be approximately 1200 (Dartmoor Commoners Council, 2016). This is a result of shrinking markets and dropping financial value, but also of the active reduction of numbers to satisfy general grazing reduction required by agri-

environment schemes introduced on Dartmoor by Natural England in the early 1990s. Furthermore, these schemes incentivise hill farmers/pony-keepers to keep cattle over ponies (Lake, 2016).

It has become important to know what genetics are being lost as these pony herds are lost. The remaining gene pool has emerged over millennia to support an invaluable source of resilient ponies.

1.4.3. Impact on Biodiversity

If ponies have a positive impact on the biodiversity of the moor, the loss of the semi-wild Dartmoor Hill Pony Herds will have a negative impact. Once lost, their adapted, inherent knowledge of how to thrive and survive the moor's harsh environment could not be reintroduced: e.g. being leared to areas on the moor, ability to prepare gorse to eat by grinding with their hooves, ability to find water in sub-zero temperatures, knowledge of 'drifting' routes when the herds are gathered in from vast areas by a limited number of people. Furthermore, it is suspected that other pony types and breeds would not have the same level of physical resilience required to survive.

There are also concerns that the loss of ponies may impact on nature conservation interests of Dartmoor. *Euphydryas aurini* are a species of butterfly which are becoming increasingly scarce in the UK. These butterflies have developed hill top behaviour and colonise when there is light cattle or horse grazing (Domino & Pullin, 2003). The Scottish natural heritage emphasises the importance of horse and cattle grazing for the survival of the species and changes in grazing regime can be detrimental to the reproduction (Scottish Natural Heritage, 2016).

There have been numerous field experiments into whether herbivores help with plant diversity. A study conducted by Lui *et al.* showed that grazing mixed herbivores together had greater plant diversity than grazing as individual species (Lui *et al.*, 2015). However, these results are conflicted with other research. Hobbs and Huenneke researched into the disruption of soil and land and compared to the diversity of the plans with grazing herbivore and areas that had not been grazed upon. They believed that herbivores damage the land and create problems for some plant growth (Hobbs & Huenneke, 1992). A review paper by Olff and Richie explains that developments have allowed explanations into these conflicts and that other studies show there is a link with herbivore grazing and plant diversity (Olff & Richie, 1998). A further study in 2004 stated that, for conservation and biodiversity of land, mixed grazing is the best option (Loucougaray *et al.*, 2004).

Gorse, which is a flowering plant ~ 3 meters tall, invades the commons of Dartmoor. The gorse invades the grounds and if left undisturbed for years it can replace long lived plant species (Rees & Hill, 2001). Previously fire was the main method of control; however, it was then discovered that fire rejuvenated gorse populations. The Hill Dartmoor consume large quantities of gorse on the common, the ponies keep the gorse population to a minimum and allow areas of the common to be accessible as the weed is unable to over populate these areas (Lake, 2016). These conservational effects could cause parts of the moor to be unreachable and contribute to extinction of species."

2. Methodology overview.

Samples were supplied from the Friends of the Dartmoor Hill Pony Society and DNA extracted using the commercially available Qiagen Blood & Tissue Kit according to manufacturer's protocol. DNA concentration and quality were assessed using a Biotek Epoch plate spectrophotometer and electrophoresis of an aliquot of each DNA sample on 1% agarose gel. All samples yielded sufficient DNA for SSR analysis, but only a subset (>75% of samples) yielded enough DNA for SNP assay as this requires greater inputs.

Two approaches to examining the genetic diversity in the populations were investigated:

1. All samples were genetically typed for a set of 15 simple sequence repeat markers drawn from the commercially available Stockmarks panel of 17 SSRs. Previous work had shown that the commercial panel, which types all 17 SSR markers in a single reaction per animal, does not perform well unless DNA quality is extremely high. We therefore divided 15 of the markers into 3 panels of 5 markers each, which has been shown to perform better across variable quality DNA.
2. A sample of animals (73 Hill Dartmoors and 14 registered Dartmoors) were tested for a panel of 172 Single Nucleotide Polymorphisms (SNPs) determined from previous studies to have a high discrimination ability between the Welsh pedigree Sections and other horse breeds. The panel is a custom made Illumina Infinium array (the "AberBeef" Chip) designed to detect and discriminate between multiple species/breeds concurrently. These 172 markers are drawn from the no longer sold Illumina Equine SNP50 assay and as such data from this analysis could be combined with existing information from Winton *et al.* (2013).

Detailed methodologies used and protocols for these tests are available on request.

2.1. Genetic analysis of Hill Dartmoors using SSR markers

2.1.1. Methodology

Using a subset of 15 of the Stockmarks equine SSR markers, B.Sc. student Nicola McElhinney amplified DNA from all 106 Hill Dartmoor samples as part of her Honours degree project. M.Sc. student Emily Ham later amplified the 16 registered Dartmoor samples supplied by Friends of the Dartmoor Hill Pony, since these arrived too close to Nicola's dissertation submission to be included in her laboratory work. Initial quality of SSR amplification was assessed by gel electrophoresis and failed samples repeated.

Samples were then analysed using capillary gel electrophoresis and allele size detection via fluorescence detection to identify the sizes of each marker product. SSR markers are variable in the size of DNA product they produce. As each animal inherits half its DNA from each parent, a marker can show two differently sized products (alleles) or, if the parents share the same alleles, a single sized product (for example, an individual might display products of 120bp and 126bp for a given SSR

marker, whilst another might be 124bp and 126bp and a third 124bp/124bp). The higher the proportion of shared allele sizes, the more closely related two animals are considered to be. Allele sizes were displayed from the fluorescence detection as peaks on a graph, and manually recorded in numerical format as above by the students.

2.1.2. Results of SSR analysis

Although overall quality of the DNA samples was good, several of the 15 markers used proved unusable due to, in one case, the marker being completely monomorphic (only one allele) across all samples (and thus useless for detecting variation) and, in another case, two markers being so similar in size that they could not be distinguished reliably. To avoid mis-scoring, these markers were excluded. Other markers worked in the Dartmoor samples but were not present in the datasets from Winton *et al.* (2013) due to similar issues. We therefore generated a combined dataset of 8 SSR markers typed across all 106 Hill Dartmoor individuals, the 16 registered Dartmoors provided by Friends of the Dartmoor Hill Pony and several other pony breeds (including 12 registered Dartmoors) collected and genotyped either as part of Winton *et al.* (2013) or for a project on the genetic relatedness of the Welsh Section A Hill Pony to the lowland Section A animals. A list of all breeds/groups used in the analysis, together with sample numbers, is summarised in Table 1.

Table 1: Breeds/groups used for SSR analysis

Group	# individuals
Sec A Welsh	52
Sec B Welsh	25
Sec C Welsh	34
Sec D Welsh	46
Carneddau	47
Connemara	47
Highland	40
Fell	39
New Forest	41
Dartmoor	28
Shetland	35
Hill Dartmoor	106
Upland Welsh	34

Once genotypes had been scored and combined into a unified dataset, this was analysed via the population genetics package GenAlEx v6.5 (Peakall & Smouse, 2012) and the STRUCTURE software v2.3 (Pritchard *et al.*, 2000). Pairwise genetic distances were calculated between each population based on F_{st} , or the variance in allele frequencies between populations, as shown in Table 2:

Table 2: Pairwise population Fst values based on SSR data

	Sec A Welsh	Sec B Welsh	Sec C Welsh	Sec D Welsh	Carneddau	Connemara	Highland	Fell	New Forest	Dartmoor	Shetland	Hill Dartmoor	Upland Welsh
Sec A Welsh	0.000												
Sec B Welsh	0.025	0.000											
Sec C Welsh	0.025	0.023	0.000										
Sec D Welsh	0.037	0.030	0.018	0.000									
Carneddau	0.023	0.031	0.037	0.039	0.000								
Connemara	0.036	0.027	0.031	0.043	0.054	0.000							
Highland	0.037	0.044	0.032	0.062	0.053	0.043	0.000						
Fell	0.042	0.034	0.024	0.035	0.050	0.029	0.032	0.000					
New Forest	0.025	0.018	0.032	0.037	0.033	0.028	0.048	0.038	0.000				
Dartmoor	0.038	0.028	0.046	0.059	0.046	0.041	0.045	0.053	0.041	0.000			
Shetland	0.046	0.052	0.051	0.064	0.054	0.075	0.079	0.073	0.050	0.092	0.000		
Hill Dartmoor	0.019	0.022	0.023	0.037	0.028	0.025	0.034	0.034	0.017	0.036	0.034	0.000	
Upland Welsh	0.049	0.040	0.054	0.066	0.048	0.046	0.052	0.065	0.059	0.055	0.072	0.039	0.000

Grey highlight shows each population tested against itself. Orange highlight shows closely related populations ($F_{st} \leq 0.02$) for both registered and Hill Dartmoors.

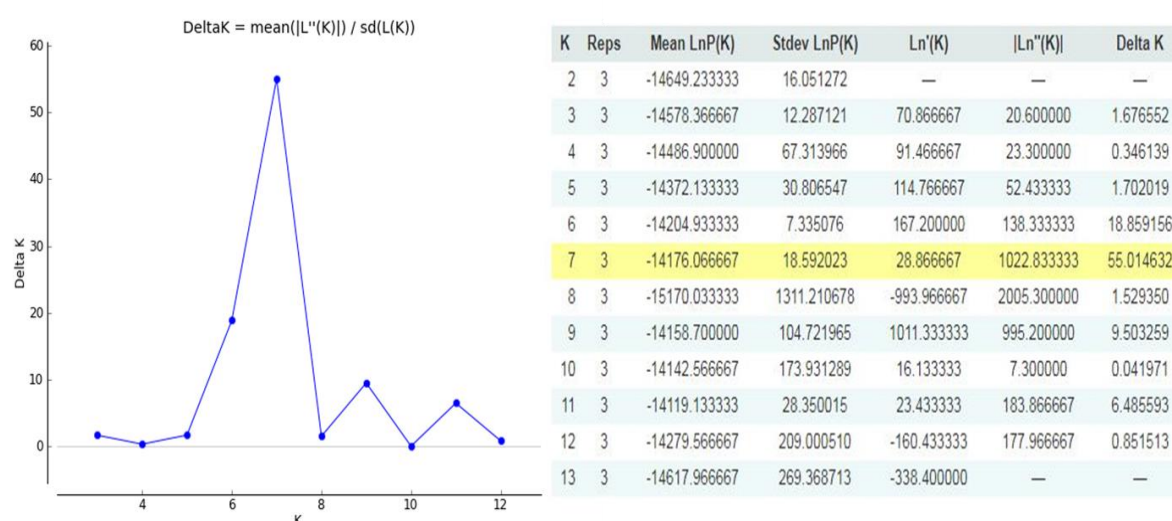
SSR marker data was then used for a population assignment test to determine whether individual animals could be successfully assigned to their population of origin, based on the alignment with the average allele frequency signature of each population, as summarised in Table 3. A more detailed results table, showing individual population assignments for each animal, is given in Appendix 1.

Table 3: Results of population assignment based on SSR data

	Sec A Welsh	Sec B Welsh	Sec C Welsh	Sec D Welsh	Carneddau	Connemara	Highland	Fell	New Forest	Dartmoor	Shetland	Hill Dartmoor	Upland Welsh
Sec A Welsh	32	5	2	1	4	1	2	2	1				2
Sec B Welsh		19		3			1			2			
Sec C Welsh	2	3	23	1	1	1		1		2			
Sec D Welsh	1	2	6	35	1					1			
Carneddau				2	43						2		2
Connemara	1	1				41		2				1	
Highland	1	1					34	3		1			
Fell	2	2		1			1	31	1	1			
New Forest		3		1		3			28	1	1	3	
Dartmoor	1	2		1	1	1		1		21			
Shetland											35		
Hill Dartmoor	9	3	3	2	3	4	6	4	6	4	7	54	1
Upland Welsh													27

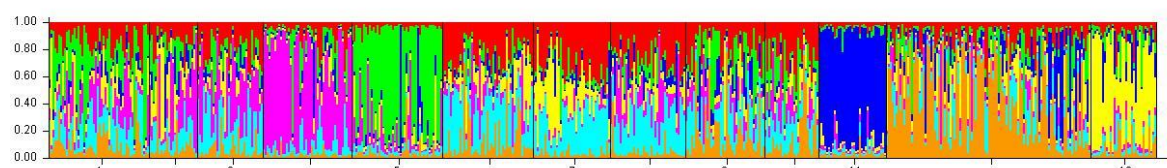
Marker data were then used for STRUCTURE analysis, where the software is asked to assume the data falls into k subgroups and then tries to assign samples to each subgroup. This provides similar information to the population assignment test above, but also shows any admixture between groups. The run parameters were set at a burnin of 25000 and 25000 replicate calculations following each burnin period. For each value of k , three replicate runs were performed. Separate runs were performed assuming that either allele frequencies were correlated with or independent of population origin – in actuality this made little difference to the observed results. Once the software had tested assumptions of k from 2 to 13 (the maximum known number of possible groups), data was analysed using STRUCTURE Harvester (Evanno *et al.*, 2005) to determine the most likely value of k .

Figure 1: STRUCTURE Harvester assessment of optimal k for SSR data



Harvester calculates an optimal value of $k = 7$ for the SSR data. A representative STRUCTURE barplot at this value is shown in Figure 2:

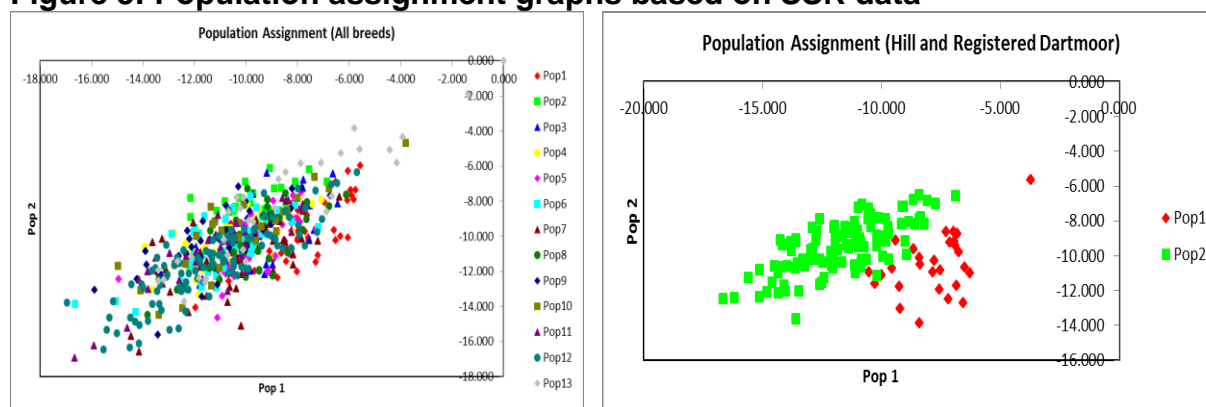
Figure 2: STRUCTURE plot for SSR data at $k = 7$



Key: 1-4 = Sec A-D Welsh, 5 = Carneddau, 6 = Connemara, 7 = Highland, 8 = Fell, 9 = New Forest, 10 = Registered Dartmoor, 11 = Shetland, 12 = Hill Dartmoor, 13 = Upland Welsh

Given that STRUCTURE analysis suggests the Hill and Registered Dartmoors can be divided into two groups, we repeated the population assignment test in GenAlEx using only the Hill and registered Dartmoor data and excluding the other breeds. Population assignment graphs for All breeds and Hill and Registered Dartmoor only are shown in Figure 3:

Figure 3: Population assignment graphs based on SSR data



Key: Lefthand graph: 1-4 = Sec A-D Welsh, 5 = Carneddau, 6 = Connemara, 7 = Highland, 8 = Fell, 9 = New Forest, 10 = Registered Dartmoor, 11 = Shetland, 12 = Hill Dartmoor, 13 = Upland Welsh. Righthand Graph: Green = Hill Dartmoor, Red = Registered Dartmoor.

Only 1 Registered Dartmoor animal was misassigned as Hill Dartmoor in this new comparison, having misassigned as Connemara in the “All Breeds” analysis. Seven Hill Dartmoors were misassigned as Registered Dartmoors. Four of these also misassigned as Registered Dartmoor in the “All Breeds” analysis but the remainder misassigned as Section A Welsh, Highland and New Forest. This suggests that the low number of markers and thus lower resolution increases the probability of random misassignment when more breeds are included.

2.2 Genetic analysis of Hill Dartmoors using SNP markers

2.2.1. Methodology

Previous work at Aberystwyth University had led to the creation of a mixed-species Illumina Infinium assay platform, the “AberBeef chip”, based on the commercially available BovineLD platform. In brief, this assay uses DNA probes bound to microbeads to interrogate sample DNA ~7000 cattle SNP markers. To detect contamination in beef samples following the HorseGate scandal, we used the Add-On Content option from the supplier to supplement the cattle SNPs with markers from horse, sheep, pig, chicken and goat and demonstrated that we could detect low levels of contamination by generating mixtures of cattle/other species DNA at known ratios. We also demonstrated that there was no cross-reactivity between probes from each species, enabling us to use the markers for genotyping experiments. The 172 SNP markers for horse were subselected from the (now discontinued) Illumina EquineSNP50 assay of ~54,000 SNPs which we had previously used in Winton *et al.* (2013) and were known to discriminate well between Welsh pony breeds and other horses.

Due to the higher DNA concentration/quality requirements compared to SSR analysis, only 73 Hill Dartmoors and 14 registered Dartmoors (all supplied by Friends of the Dartmoor Hill Pony) were suitable for testing, but this was deemed sufficient to compare to existing data from Winton *et al.* (2013). Sample DNA was mixed with cattle DNA (from a separate project with the Hereford Original Population Society) at a ratio of 80:20 to generate enough signal from the cattle markers to scan correctly. Samples were bound to the assay chips as per the manufacturer’s protocol (https://support.illumina.com/documents/myillumina/67f59f89-51ee-44d6-b1bb-a53dcb5bd01e/infinium_hd_ultra_user_guide_11328087_revb.pdf) and chips scanned using the Illumina HiScanSQ platform in Dr. Hegarty’s laboratory. Briefly, for each marker, allele 1 is labelled with a green dye, allele 2 with a red dye. If the individual is homozygous for allele 1 for a given SNP, the microbeads for that SNP will show a green signal, and red if homozygous for allele 2. If the individual is heterozygous there will be both green and red signal which the scanner sees as yellow. Following scanning and recording of colour signals for all SNPs on the array, the data is transferred to the Illumina GenomeStudio software and converted from colour signal data to human-readable genotypes (AA if homozygous for allele 1, BB if homozygous for allele 2 and AB if heterozygous).

Because GenAIEx and STRUCTURE do not accept this format, the data were further processed in Excel to split the genotype into two columns for each marker and then recoded so A = 1, B = 2. Hence, a homozygous individual would be either 11 or 22,

with heterozygotes being 12. Missing data (where genotyping had failed for a given marker/sample combination) was rare but such cases were assigned values of -9 as recognised by GenAlEx and STRUCTURE both. The data were then combined with similarly-treated SNP data derived from Winton *et al.* (2013).

2.2.2. Results of SNP analysis

All samples appeared to genotype successfully given the Illumina QC parameters (allele call rates of 95% or above). Upon entering the data into GenAlEx, some issues were identified with high levels of missing data for 4 markers and 1 sample, necessitating their removal prior to further analysis. This is common to these types of analysis as some markers do not genotype well across all populations and some samples may be of threshold DNA quality for genotyping. The combined dataset therefore contained genotypes for the following breeds/populations across a total of 168 SNP markers, as per Table 4:

Table 4: Populations used for SNP analysis

Group	# Individuals
Sec A Welsh	12
Sec D Welsh	12
Dartmoor	14
Hill Dartmoor	73
Irish Draught	8
Connemara	12
Carneddau	12
Upland Welsh	67

As with the SSR data, GenAlEx was used to assess the pairwise F_{st} as a measure of genetic distance between populations (Table 5):

Table 5: Pairwise population F_{st} values based on SNP data

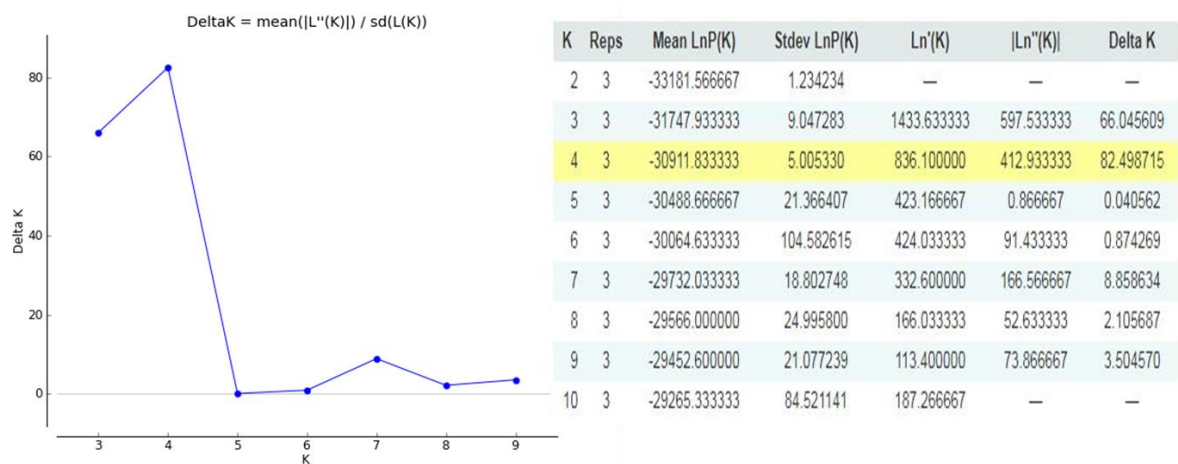
	Sec A Welsh	Sec D Welsh	Dartmoor	Hill Dartmoor	Irish Draught	Connemara	Carneddau	Upland Welsh
Sec A Welsh	0.000							
Sec D Welsh	0.333	0.000						
Dartmoor	0.151	0.193	0.000					
Hill Dartmoor	0.121	0.175	0.024	0.000				
Irish Draught	0.198	0.208	0.069	0.054	0.000			
Connemara	0.277	0.280	0.126	0.105	0.104	0.000		
Carneddau	0.157	0.318	0.093	0.078	0.142	0.220	0.000	
Upland Welsh	0.039	0.229	0.079	0.054	0.121	0.192	0.097	0.000

Population assignment testing was then undertaken in GenAlEx as for the SSR data:

Table 6: Results of population assignment based on SNP data

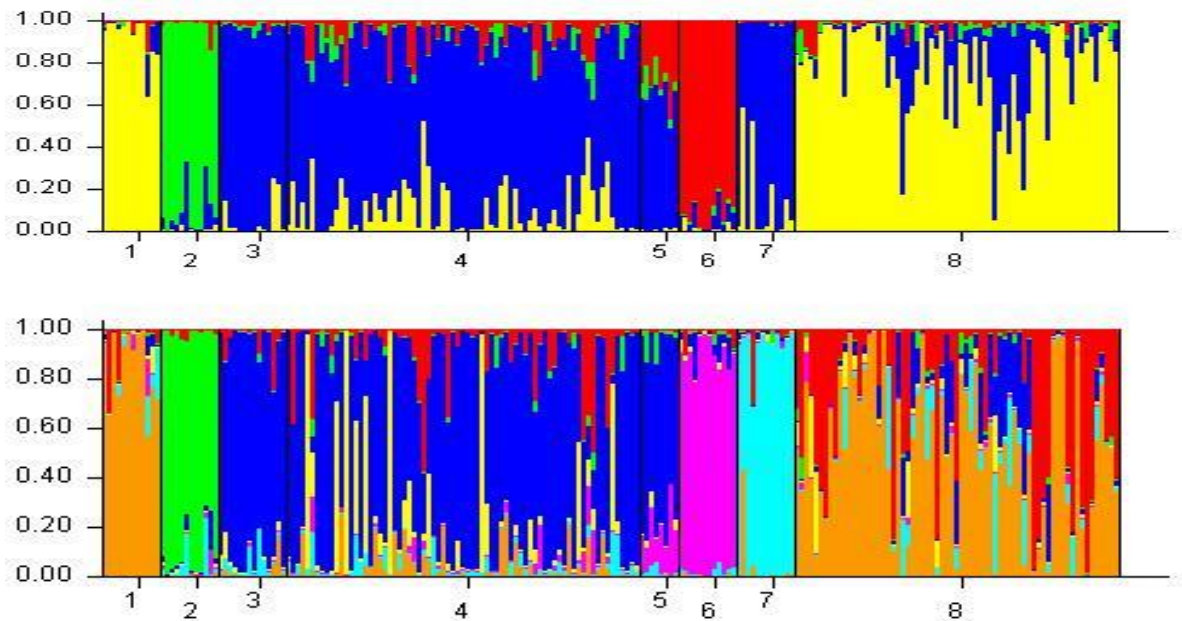
	Sec A Welsh	Sec D Welsh	Dartmoor	Hill Dartmoor	Irish Draught	Connemara	Carneddau	Upland Welsh
Sec A Welsh	12							
Sec D Welsh		12						
Dartmoor			14					
Hill Dartmoor			1	71				1
Irish Draught					8			
Connemara						12		
Carneddau							11	1
Upland Welsh	1			3				63

STRUCTURE analysis was then performed using the same parameters as for the SSR data, except values of k from 2 to 10 were tested given the smaller number of known populations used.

Figure 4: STRUCTURE Harvester assessment of optimal k for SNP data

Optimal k here was determined to be 4, but this is likely to be detecting ancestral linkages due to the slower evolution of SNP markers compared to SSRs. A secondary peak at $k=7$ was detected which agrees with the SSR result of $k=7$. Barplots for both values of k are shown in Figure 5:

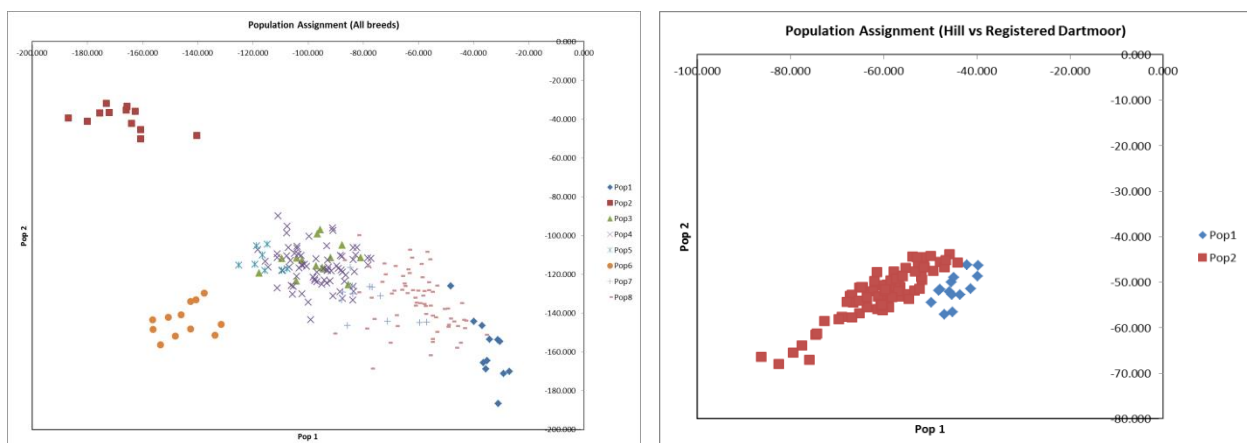
Figure 5: STRUCTURE plot for SNP data at $k = 4$ (top) and $k = 7$ (bottom)



Key: 1 = Sec A Welsh, 2 = Sec D Welsh, 3 = Registered Dartmoor, 4 = Hill Dartmoor, 5 = Irish Draught, 6 = Connemara, 7 = Carneddau, 8 = Upland Welsh

As before, once it was determined that STRUCTURE supported the division of Hill and Registered Dartmoors into two groups, we repeated the population assignment test in GenAlEx using only the Hill and registered Dartmoor data and excluding the other breeds. Population assignment graphs for All breeds and Hill and Registered Dartmoor only are shown in Figure 6:

Figure 6: Population assignment graphs based on SNP data



Key: Lefthand graph: 1 = Sec A Welsh, 2 = Sec D Welsh, 3 = Registered Dartmoor, 4 = Hill Dartmoor, 5 = Irish Draught, 6 = Connemara, 7 = Carneddau, 8 = Upland Welsh. Righthand Graph: Red = Hill Dartmoor, Blue = Registered Dartmoor.

In the Hill vs Registered analysis, only a single Hill Dartmoor individual was misassigned as Registered Dartmoor – this individual was known to be the offspring of a Registered X Hill cross. In the “All Breeds” analysis, only one other Hill Dartmoor sample was misassigned (as Upland Welsh).

Given the higher ability of SNP markers to distinguish between populations, as shown by these results and in the lefthand graph of Figure 6, we then attempted to determine which markers were strongly distinguishing between the Hill and Registered Dartmoors. To do this, the SNP data for all breeds were converted into PLINK format (Purcell *et al.* 2007) and loaded into the gPLINK software package. Samples were assigned as “case” (Hill), “control” (Registered) or “unknown” (all other breeds) and a genetic association test undertaken. Briefly, this identifies SNP markers showing a significant difference in allele frequency between “case” and “control” groups, whilst taking the “unknowns” into account. Therefore, the test only identifies markers which still show differences between the Hill and Registered Dartmoors after excluding all of the variation present in other breeds. The test identified five SNP markers displaying significant ($p < 0.05$ after Bonferroni multiple testing correction) differences in allele frequency. A further two markers were significant at $p < 0.1$ and were also investigated. As the positions of these markers on the *E. caballus* genome are known, we investigated their location relative to known genes to determine if there may be any link to phenotypic (trait) differences between the breeds. The markers are listed below:

Table 7: SNP markers significantly varying in allele frequency between Hill and Registered Dartmoor groups

SNP	Freq (Hill)	Freq (Reg)	P	BONF	Located in/near gene(s)
BIEC2_1138852	0.09028	0.5	6.69E-10	1.07E-07	Near IL1RAPL2, TEX13A
BIEC2_670047	0.09722	0.3913	3.49E-06	0.000558	In NTMT1, near ASB6
BIEC2_729317	0.2361	0.5435	8.77E-05	0.01404	Near SLC6A15
BIEC2_346872	0.2014	0.4783	0.000225	0.036	In SLC22A14, near XYLB
BIEC2_346873	0.2014	0.4783	0.000225	0.036	In SLC22A14, near XYLB
BIEC2_1134509	0.2847	0.5652	0.000528	0.08441	None identified
BIEC2_703378	0.4028	0.6957	0.000534	0.08536	Near UNC5D

Key: SNP = Marker name as per International Equine Consortium, Freq = frequency of major allele (1 = 100%), P = probability of significance (1 = not significant), BONF = Bonferroni-corrected P value for multiple testing

3. Discussion

3.1. Genetic relatedness of Hill Dartmoor to other pony breeds

Pairwise F_{st} analysis of the populations showed that, on the basis of SSR markers, Hill Dartmoors are genetically closer to several of the other pony breeds (particularly Section A Welsh, New Forest and Highland) than they are to Registered Dartmoors. However, SNP analysis shows the two groups to be much closer than to others. This conflict in results is likely due to the lower number of SSR markers used and thus lower power to accurately distinguish between groups. This is evidenced by the population assignment test for SSR markers which shows that, while for the most part animals can be assigned to the correct group, there are significant numbers of misassigned animals in most breeds (except Shetlands, a highly isolated breed). Hill Dartmoors are particularly prone to misassignment with SSR markers when compared to other breeds, notably to other upland types. A number of Hill samples are also the product of Hill X Shetland crosses and these animals typically assign as Shetland. There are relatively few SSR alleles which are unique to a particular breed and, as a free-living population, the Hill ponies will therefore be segregating for a

wider range of alleles than in the closed breed populations, contributing to higher genetic diversity but also an increased chance of misassignment when using low numbers of genetic markers.

Using a higher number of SNP markers we see in Figure 6 that there is much greater power to discriminate populations than observed in Figure 3 for SSRs. Pairwise F_{st} values here show the Hill and Registered Dartmoors to be more closely related to each other than to other groups, as we would expect given their history. We observe similar (though smaller) genetic distances to those seen between the Section A Welsh pony and Upland Welsh populations. Indeed, the Upland Welsh and Carneddau also appear genetically close to both the Hill and Registered Dartmoor (more so the Hill), confirming our preliminary findings and hinting at possible shared adaptations to hill living, though further work will be needed to confirm this.

3.2. Population genetic components

STRUCTURE analysis of the SSR and SNP data shows that, in both cases, differences in assignment to genetic clusters can be observed between the Hill and Registered Dartmoors. This is more apparent when using SSR markers, as per Figure 2, where the Hills show a much greater proportion of the orange component seen in the background of several upland types. Towards the last third of the Hill set can also be seen the influence of the Hill X Shetland crosses, which display the blue component identified as the main signature of the Shetlands. The SNP plot, as with the pairwise F_{st} analysis, shows a closer relationship between the Hill and Registered samples but, as seen in Figure 5, the Hills display more of the same component as seen in the Upland Welsh (yellow in the upper plot, red in the lower) and, at $k = 7$, begin to display a yellow component not typically observed in any of the other groups. At the lower value of $k = 4$, the Carneddau population show a high degree of similarity to both the Hill and Registered Dartmoors, but appears clearly distinct as $k = 7$, suggesting that this value is more accurate. In both SSR and SNP analyses at $k = 7$, the Hills do form a group more similar to themselves (though highly diverse) than to others. This is borne out when we subsequently repeated the GenAlEx population assignment testing based solely on the Hill and Registered Dartmoor samples – the SNP analysis had previously proven very robust in distinguishing the two groups even in the presence of the other breeds, but now the SSR analysis is much more capable of distinguishing the two groups, with only one Registered Dartmoor being misassigned and 7 Hill Dartmoors (including one known to be part of a cohort using a Registered stallion).

3.3. Candidate genes under divergent selection in the Hill and Registered Dartmoor ponies

Using PLINK analysis of the SNP markers, we identified seven SNPs showing significant differences in allele frequency between the Hill and Registered Dartmoor groups (Table 7). The most significant marker (BIEC2-1138852) shows a dramatic shift in the frequency of the most common (major) allele – as represented across all breeds – when comparing just the Hill and Registered Dartmoors. In the Registered Dartmoor, the major allele is present in 50% of the animals tested, whereas it is only present in 9% of the Hill animals. This pattern would be typical of a marker which is under selection in the Hills but not the Registered Dartmoors – without selection,

there would be no consequence of having either allele and both would exist in an ~50:50 ratio. A strong shift in allele frequency is typically the result of small founder populations (more likely in the Registered ponies than the Hills due to their history), human selection (again, more likely in the Registered ponies) or natural selection. The remaining markers show similar patterns of reduced frequency of the major allele in the Hills, varying from 20-60% of the frequency seen in the Registered animals. The genetic positions of these on the equine reference genome were used to identify which gene(s) the SNPs lay within or near – due to the principles of genetic linkage, even if the marker itself is not under selection, if it lies close to a gene that is, it will show associated changes in allele frequency and can thus be used as a diagnostic of that gene. All but one of the markers lay in/near genes, which were then reviewed in the scientific literature to identify any possible links to known trait differences between Hill and Registered Dartmoors.

The most significant marker (BIEC2-1138852) lies on the X chromosome near genes IL1RAPL2 and TEX13A. IL1RAPL2 is an interleukin-like receptor-associated protein which has been shown to be involved in autoinflammatory conditions (i.e. arthritis) and general inflammation response as part of the immune system (GeneCards). Intriguingly, it has been associated with cold hypersensitivity in humans, whereby exposure to cold causes an inflammatory response (GeneCards). Given the upland environment of the Hill Dartmoors, a possible role for this gene in cold tolerance is worth investigating further. TEX13A is expressed exclusively during sperm development in the testes, where it controls the activity of ~130 other genes. It is not known if there are differences in male fertility between the Hill and Registered animals, although fertility issues should be selected against in both cases.

Marker BIEC2-670047 lies within the NTMT1 gene on chromosome 25 and close to the gene ASB6. This genomic region has been identified through genome-wide association analysis of pigs as being linked to backfat thickness (Casiro, 2016). ASB6 has been identified as playing a role in abdominal fat in chickens (Liu *et al.*, 2013). The Hill Dartmoors are known for their ability to retain fat during the harsh weather and food shortage of winter months, so there is a strong possibility that one or both of these genes is linked to that trait. The NTMT1 gene is better-known for its role in DNA damage repair and cancer prevention/control of cell growth. Whilst the link to fattening/overall body growth rate is easier to explain phenotypically, it is also possible that the Hill ponies are exposed to higher levels of UV light and thus DNA damage.

Three of the markers lie in/near solute carrier proteins (SLCs), although two of these are actually within the same gene (lying 760bp apart). SLC6A15 is a neuronal transmitter which has been strongly associated with depression and anxiety in humans and other mammals. Differential responses to stress may be a necessary adaptation to upland living, although further work with a trained animal behaviourist would be needed to confirm this. SLC22A14 is again known to be associated with body fat composition in pigs (Reyer *et al.*, 2017). The markers in this gene are also close to the XYLB gene, implicated in pulmonary hypertension.

The final marker lies near the gene UNC5D on chromosome 27. This gene is linked to ADHD and Alzheimer's in humans, indicating a possible role again in behavioural response. There is also a link to the effects of maternal obesity on embryo

development (Sanders *et al.*, 2014), which is interesting given the ability of the Hills to retain fat. Interestingly, this response is also linked to the interleukin receptor pathway, perhaps suggesting a link to the IL1RAPL2 hit.

4. Conclusions

Returning to the questions raised by the Friends of the Dartmoor Hill Pony at the start of this project:

i) Are Dartmoor Hill Ponies one distinct genetic type of pony? To what degree are they distinct from, or similar to, other populations of native pony breeds? To what degree are they distinct from their nearest cousin, the registered Dartmoor Pony?

Population assignment analysis via GenAIEx shows that the bulk of Hill Dartmoors can be assigned to a single group, though this is less certain with SSR markers due to the lower resolution of the analysis. When using SNP markers, certainty increases to >97%. In both types of analysis, confidence improves further when comparing solely the Hill and Registered Dartmoors. STRUCTURE analysis confirms that the two groups can be separated, showing an increased proportion of genetic signatures seen in other upland pony types compared to the Registered animals. The SNP analysis should be considered the gold standard here due to the higher resolution and suggests that the Hill Dartmoors are related to the Registered breed but contain an additional genetic component more similar to the Upland Welsh, as well as (at $k = 7$) a novel genetic signature of their own. The pairwise F_{st} value for SNP data of 0.024 between the Hill and Registered groups is considerably closer than to any of the other breeds (though the Upland Welsh and Carneddau remain close, hinting at shared ancestry or adaptation to hill living), confirming the results of the population assignment/STRUCTURE testing. This value could be compared to the pairwise F_{st} between humans from an isolated region (Finland) compared to the rest of Europe and is $\sim 1/5^{\text{th}}$ the pairwise F_{st} difference between European and Asian humans (Nelis *et al.*, 2009).

ii) As the number of Dartmoor Hill Ponies decreases, are we losing valuable genetics?

There is evidence of a genetic component in the Hill Dartmoors which is shared with other upland types and which is not present (or reduced) in the Registered Dartmoor. Without further study, it is not possible to say for certain whether this component is “valuable” (although from a population perspective a higher degree of genetic diversity is always preferable to allow adaptation to a wider range of selective pressures). PLINK analysis identified a set of 7 SNP markers (out of 168 tested) showing significant differences in allele frequency between the two groups. In the most extreme case, the Hills favour a rarer allele not seen in the majority of ponies and present in 91% of Hills versus 50% of Registered Dartmoors. This suggests that the rarer allele is under selection in the Hill Dartmoors and, given the potential link to cold tolerance, may represent an adaptation to hill environments. The other markers also favour rarer alleles in the Hills, although to a lesser degree. In several of these cases, there are potential links to known phenotypes in the Hill pony (ability to gain/lose fat). Breeding the Hill Dartmoor with the Registered Dartmoor will have the

potential to remove these rarer alleles if they are not under selection, with a 25-50% chance of losing the rarer allele at each generation of backcrossing depending on whether the other parent is heterozygous or homozygous for the more common allele. The layperson's viewpoint may be that the Registered Dartmoor still has the rarer allele in 50% of cases, but from a genetics perspective it is important to understand that genes operate in networks and it is the overall makeup that contributes to an organism's ability to adapt to its environment. Without the selective pressure, alleles may be retained but the adaptive combination of alleles will break down. The high proportion of the rarer allele for the marker discussed above, for example, means that almost all Hill Dartmoors will be homozygous for that allele – indicating a recessive trait that is favoured by selection. Individuals carrying only one copy of the allele would not demonstrate the same adaptation.

To conclude, whilst the Hill and Registered Dartmoors are closely related, there are clear genetic differences which enable the groups to be distinguished, as well as signs of genetic selection potentially linked to environmental adaptation, although further research would be required to confirm this.

5. Literature Cited

Bowling. A. (2001). Historical development and application of molecular genetic tests for horse identification and parentage control. *Livestock Production Science*. 72 (1-2), 111-116.

Casiro S (2016) GENOME-WIDE ASSOCIATION STUDY IN AN F2 DUROC X PIETRAIN RESOURCE POPULATION FOR MEAT QUALITY AND CARCASS TRAITS. Ph.D thesis (University of Michigan)

Dartmoor Pony Society (2016) About the Breed. Available <http://www.dartmoorponysociety.com/aboutthebreed.html>. Last accessed 1st May 2017.

Dartmoor National Park Authority (2015) The Dartmoor Ponies. Available: <http://www.dartmoor.gov.uk/learningabout/lab-printableresources/lab-factsheetshome/lab-dartmoorponies>. Last accessed 2nd May 2017.

Domino J. and Pullin AS (2003). Conservation implications of the distribution of genetic diversity at different scales: a case study using the marsh fritillary butterfly (*Euphydryas aurinia*). *Biological Conservation*. 114 (3), 453–461.

Evanno et al., 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* 14, 2611–2620.

Friends of the Hill Dartmoor (2013) History. Available: <http://www.friendsofthedartmoorhillpony.co.uk/history.html>. Last accessed 2nd May 2017.

Genecards. <http://www.genecards.org/cgi-bin/carddisp.pl?gene=IL1RAPL2#diseases> (accessed 19/12/2017)

Hobbs.R.J and Huenneke. L.F. (1992). Disturbance, Diversity, and Invasion: Implications for Conservation. *Conservation Biology*. 6 (3), 324-337.

Lake, S. (2016) Upland Pony grazing: a review. Footprint Ecology/Dartmoor's Pony Action Group.2

Liu R, Sun Y, Zhao G, Wang F, Wu D, Zheng M, *et al.* (2013) Genome-Wide Association Study Identifies Loci and Candidate Genes for Body Composition and Meat Quality Traits in Beijing-You Chickens. *PLoS ONE* 8(4): e61172. <https://doi.org/10.1371/journal.pone.0061172>

Loucougaray, G., Bonis, A. & Bouzillé, J.B. (2004) Effects of grazing by horses and/or cattle on the diversity of coastal grasslands in western France. *Biological Conservation*, 116(1), pp.59–71.

Nelis M, Esko T, Mägi R, Zimprich F, Zimprich A, *et al.* (2009) Genetic Structure of Europeans: A View from the North–East. *PLoS One*, 4(5): e5472. <https://doi.org/10.1371/journal.pone.0005472>

Olf H & Richie ME (1998) Effects of herbivores on grassland plant diversity. *Trends Ecol Evol.* 13(7): 261-265.

Palmer, J. (1990). *The Dartmoor Pony: A History of the Breed*. Exeter: Devon Books
Pritchard.J.K, Stephens.M, and Donnelly.P. (2000). Inference of Population Structure Using Multilocus Genotype Data. *Genetics.* 155 (2), 945-959.

Peakall R & Smouse PE (2012) GenAEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics* 28(19): 2537–2539.

Pritchard J, Stephens M & Donnelly P (2000) Inference of Population Structure Using Multilocus Genotype Data. *Genetics* 155 (2) 945-959.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ & Sham PC (2007) PLINK: a toolset for whole-genome association and population-based linkage analysis. *American Journal of Human Genetics*, 81.

Rees M & Hill RL (2001) Large-scale disturbances, biological control and the dynamics of gorse populations. *J. Applied Ecology* 2: 364–377.

Reyer H, Varley PF, Murani E, Ponsuksili S & Wimmers K (2017) Genetics of body fat mass and related traits in a pig population selected for leanness. *Scientific Reports*, doi:10.1038/s41598-017-08961-4

Sanders TR, Kim DW, Glendining KA, Jasoni CL (2014) Maternal obesity and IL-6 lead to aberrant developmental gene expression and deregulated neurite growth in the fetal arcuate nucleus. *Endocrinology* 155(7):2566-77.

Scottish Natural Heritage (2016), Marsh fritillary (*Euphydryas autinia*). Scottish Natural Heritage, Inverness.

The Dartmoor Hill Pony Association (2015) About Dartmoor Hill Ponies. Available: <http://www.dartmoorhillpony.com/>. Last accessed 1st May 2017.

Winton, C. L., Hegarty. M.J, McMahon. R, Slavov. G.T, McEwan. N.R, Davies Morel. M.C.G, Morgan. C.M, Powel. W, and Nash. D.M. 2013. Genetic diversity and phylogenetic analysis of native mountain ponies of Britain and Ireland reveal a novel rare population. *Ecol. Evol.* 3:934–947.